Towards a First Generation of Hourly Animations of Boundary Layer Heights along the East Coast *Streaming* from a Lidar-based Ground Observing System

Presentation: M.Halem, University of Maryland, Baltimore County

Faculty:

CSEE: P. Nguyen, J. Sleeman, D. Chapman, J. Dorband, Ya. Yesha

JCET: B. Demoz, R. Delgado, V. Caicedo,

Students: Z.Yang*, R. Gite**, S. Shivadekar*, P. Bindu**, K. Patel**,

Z. Ali**, D. Ziaei***

(A UMBC collaboration between CSEE and JCET)

NNH16ZDA001 AIST-16-0091

Supplement P0011 (12/1/20-11/30/21)

halem@umbc.edu * PhD or Pre_PhD Defense/scheduled ** MS Defense/scheduled ***PhD Award

ESTO Earth Science Technology Office

Demonstration Using Streaming, AI/ML and Visualization for a Lidar Observed Aerosol Boundary Layer Height (ablh) as an Hourly Product Over the East Coast

- What is ablh? Aerosols are particulate matter (solid, liquid) in air of ~1-10 nm size (dust, carbon, pollen, sea salt, NOx, SO_2 etc.) and ablh is a transition surface between aerosols in pbl and the free atmosphere.
- Why do we need it?

  Science: J. Hansen- largest remaining model uncertainty for global warming; mostly due to uncertainty of fraction of aerosol in the current atmosphere. Spread of uncertainty has not changed in 40 years since Charney CO2 Academy Report.
  H. Morrison- Microphysics modeling of clouds and precipitation is key link between water and energy cycles posing the leading uncertainty in NWP and CP. Recent approaches spectral binning, LES and particle schemes such as Lagrangian super particles, though extremely computational costly offer potential convergence to turbulence and uncertainty resolutions.
  R. Kahn etal,- Relationship between pblh and surface pollutants, especially particulate matter (PM) concentration, is not yet well understood.

  Health: "60 percent of Americans live in areas where air pollution has reached unhealthy levels that can make people sick".

  Disasters: Impacts of pollutants can arise from wildfires, volcanoes, dust storms as in 30s, 50s, etc and climate change.

- Related Works.

  EPA AirNow- hourly maps of air/surface pm2.5 and 03 from prior extrapolated observations but no ablh.

  NOAA/EPA - developed a national Air Quality Forecast (AQF) System to provide ozone, particulate matter, and other pollutant forecasts.
  IBM GRAF - model issues 3 hr global Air Quality forecast. Not sure if they observe or assimilate ablh.

- Preliminary Results: Scaled end-to-end, edge streaming boundary layer height hourly product.

- Contributions: First 15s streaming aerosols profiles from distributed locations and instruments; processed ablh using AI/ML to avoid potential of small noisy derivatives; animation by fusing with GFS.

Task 1: NASA/AIST/UMBC Grant Ceilometer Deployment for this Demonstration

- **Locations**

1 - Pennsylvania Department of Environmental Protection
Bristol PA Air Quality Monitoring Station

2 - **UMBC** with Ceilometer intended for Navajo Technical University
Integration of Remote Sensing to Computer/Environmental Science Dept

3 – **UMBC** with Ceilometer borrowed from Essex MD

4 – **UMBC/ JCET/EPA Catonsville** ceilometer on top of Physics Bldg.

5 - **Virginia Tech** (Elena Lind) ? **Blacksburg Va.** on top of Environmental Science CVenter
Bradley Dept Electrical and Computer Engineering
Ceilometer Aerosol Profiling (PBLH) to aid PANDORA profiling retrievals

https://www.timeanddate.com/weather/@5181688/historic?month=9&year=2020

xAce Hardware Infrastructure

D. Chapman and J.Dorband

Infrastructure for Aerosol processing

Claude 1&2 servers ($4K)

- Dual 14.2 Teraflop Nvidia Geforce 2080Ti CUDA capable GPUs (~Nvidia V100)
- 32 Core AMD Ryzen Threadripper 2990wx (~Epyc) 3.0 GHz CPu

Drobo storage configuration ($5K)

- 96 Terabyte Network Attached Storage

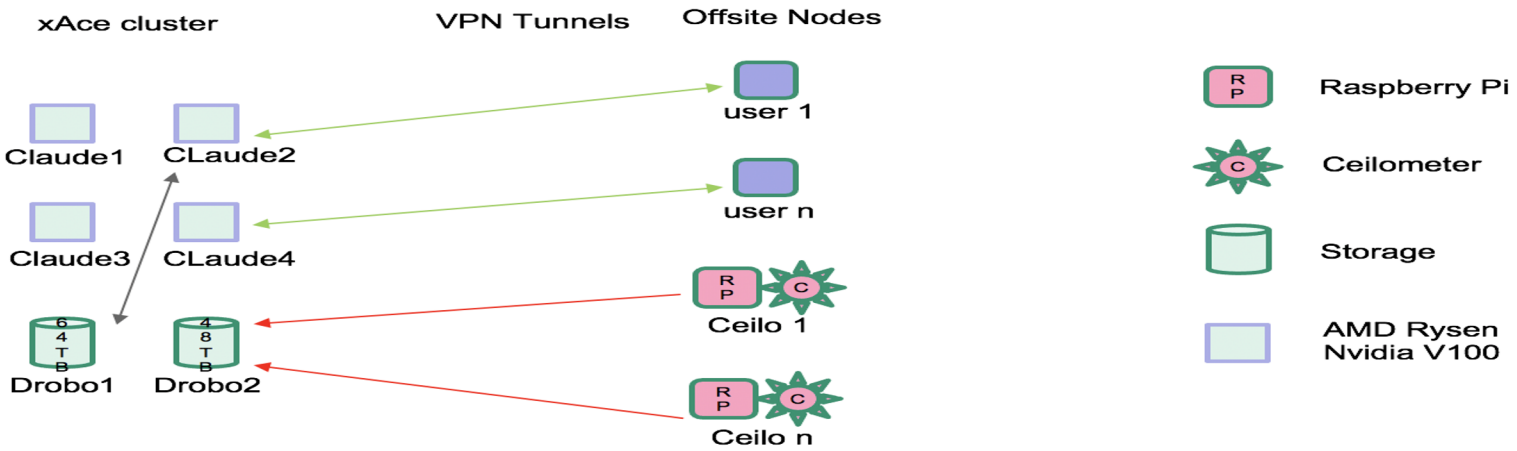
PFSense Firewall with VPN ($250)

Raspberry Pi + Ethernet Adapters (4 @ $50 Ea.)

xAce Claude and Drobo servers

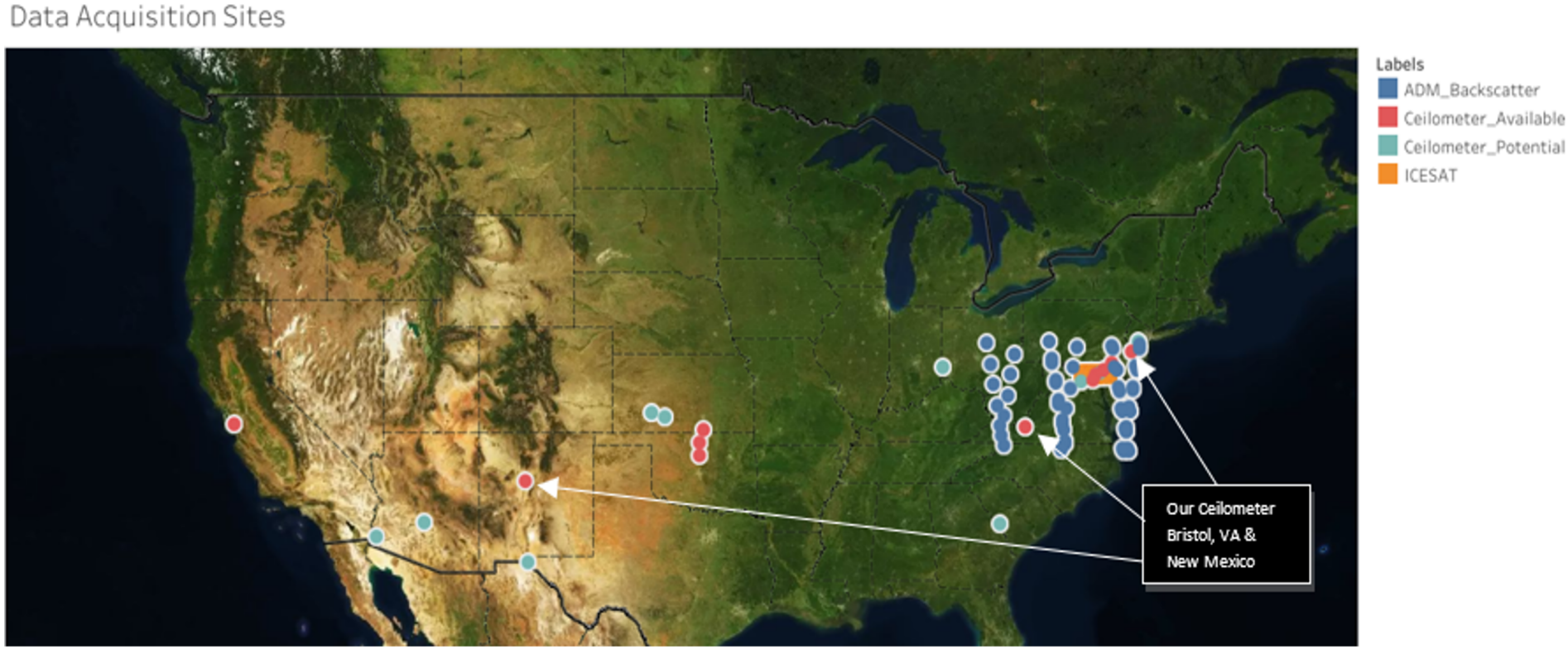
Aerosol Cyberinfrastructure Ecosystem (xACE)
J. Dorband

- Intra-net Security
  - VPN access security (user unique certificate & password)
  - Node security (VPN access & user unique password)
  - Connections in are secure
  - Connections out are open
- Once connected to Intra-net:
  - Access from any machine to any other machine with valid user account
    - User workstation (laptop, desktop)
    - Compute nodes
    - Instrument node
- Instrument node (~ $50 Raspberry Pi)
  - Local data backup from instrument (up to 3 yrs)
  - Periodically passes data on to xAce cluster database
  - Can send data to other offsite nodes/organizations (future)



Task 1. Aggregate Acquisition Processing System
P. Nguyen, R. Gite, S. Shivadekar

Task 1: Identify, negotiate, acquire and implement an internet based distributed edge streaming computing system of Level 1 ground-based ceilometer LIDAR PBLH observations over the CONUS.



Data source: Ceilometers/Radiosondes from AIST/CSEE grant, UMBC/JCET, DOE/ARM, San Jose University
Model Output: from NOAA (HRRR hourly 3km), our WRF model output.
Satellites IceSat-2 backscatter, ADM wind, backscatter radiation

Ceilometers and Model output

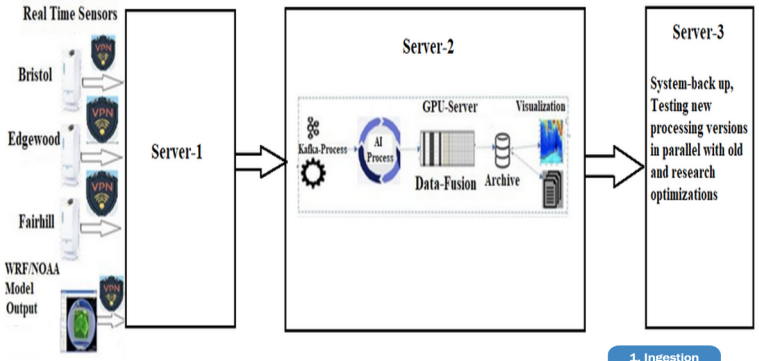
- Ceilometers: 3 Ceilometers from our grant at Bristol and VA Tech and NTU
  4 Ceilometers from JCET UMBC (1 ceil ~4 years of data from UMBC)
  3 Ceilometers from ARM SGP
- Validation: Field campaign PECAN Ceilometers, 3 Radiosondes (near UMBC, VA tech, Bristol PA)
- Satellites: NASA Icesat-2, potential ESA Aeolus ADM data products
- Model Output: NOAA (HRRR hourly at 3km), WRF-CHEM-GOCART

Station	Latitude	Longitude	Location	DATA Source	Start Date(YYYY-MM-DD)	End Date(YYYY	Total Data Size
ARM/Southern Great Plains C1	36.605	-97.485	LAMONT, OK (Permanent)	ARM	2000-05-22	2020-09-10	102 GB
ARM/Southern Great Plains E9	37.133	-97.266	ASHTON, KS (EXTENDED)	ARM	2020-01-17	2020-09-10	546 MB
ARM/Southern Great Plains E36	36.1166	-97.5112	(EXTENDED)	ARM	2020-01-17	2020-09-10	318 MB
San José State University (SJSU)	37.3327	-121.882	CA	SJSU	2019-04-10	2020-09-10	6.35 GB
University of Maryland, Baltimore County (UMBC)	39.255	-76.7095	Baltimore County,MD	UMBC	2016-12-01	2020-09-10	23 GB
Howard University Beltsville (HUB)	39.0553	-76.8783	Beltsville, MD	UMBC	2020-02-02	2020-09-10	3.7 GB
Bristol	40.1007	-74.8518	Bristol, PA	UMBC	2020-06-30	2020-09-10	0.8 GB
Virginia Tech, Blacksburg	37.2296	-80.4139	Blacksburg, VA	UMBC	2020-08-05	2020-09-10	0.8 GB
New York City	40.7128	-74.006	New York City, NY	UMBC	2020-07-01	2020-09-10	2 GB
Fair Hill (FAIR)	39.7014	-75.8601	Fair Hill,MD	UMBC	2020-02-01	2020-09-10	2.5 GB
Edgewood (EDGE)	39.4102	-76.2969	Edgewood, MD	UMBC	2020-02-01	2020-09-10	4.28 GB
PECAN campaign ceilometer	6	sites	Multiple location		2015-05-01	2015-07-30	5GB
The City College of New York (CCNY)	40.8202	-73.9503	New York City	UMBC			
WRFOutput_MYNN			US Conus	UMBC	2020-01-25	2020-01-31	101 GB
WRFOutput_YSU			US Conus	UMBC	2020-01-25	2020-01-31	90 GB
NOAA_Output			US Conus	NOAA	2020-03-09	Current Date	
WRF_Chem			US Conus	UMBC	2018-01-01	2019-06-08	249 GB
ARM SGP Radiosondes	36.6	-97.49		ARM	2019-04-10	2020-02-21	0.188GB
PECAN campaign Radiosondes	5	sites	Multiple location				
ATLAS/ICESat-2 L3A	39.168	39.3382	UMBC	NASA	2018-10-13	2020-05-24	158 GB
ATLAS/ICESat-2 L3A	36.615	36.605	LAMONT, OK	NASA	2018-10-13	2020-05-24	38.9 GB
ATLAS/ICESat-2 L3A	36.1166	36.1176	MARSHALL, OK	NASA	2020-03-17	2020-05-24	4.73 GB
ATLAS/ICESat-2 L3A	37.143	37.133	ASHTON, KS	NASA	2020-03-17	2020-05-24	4.68 GB
							Total ~ 1TB

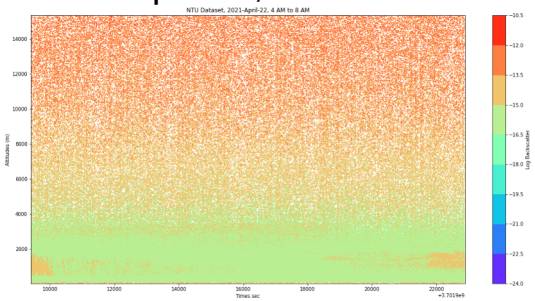
Edge Streaming Prototype: Secure, Fault Tolerant, Auto Ingestion, Processing and Analytics

P. Nguyen,S. Shivadekar, R. Gite

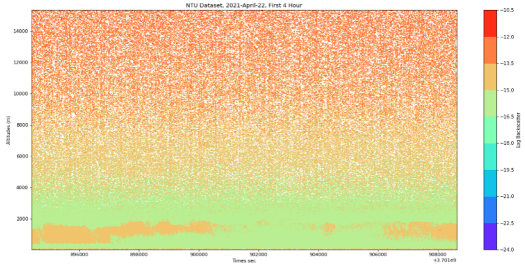
- **Built Ingestion Server:** use Apache Kafka to handle streams of data from multiple ceilometers automatically and backing up pre-processing (raw ceilometer backscatter profiles) Level 1B daily data products.
- **Pulling data** from NOAA's GSF Model Output (PBLH, HRRR hourly product) and 3 ceilometers data sites from ARM SGP (automatically) via HTTP/FTP/Web service API.
- Kafka Server starts and runs *our developed Ingestion software component connects to instrument end point (via Ceilometer-*RaspberryPi-*Claude Server) ~4MB connection.*
- Ingestion keeps checking on latest update files→ Pulling updated files to Claude server → Save file to server at Drobo storage location→ 4. Process pipeline
- Video show streaming from NTU instrument's observation to Claude server

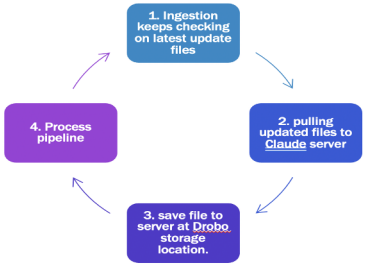


NTU April 22, 2021 0-4AM



NTU April 22, 2021 4-8AM

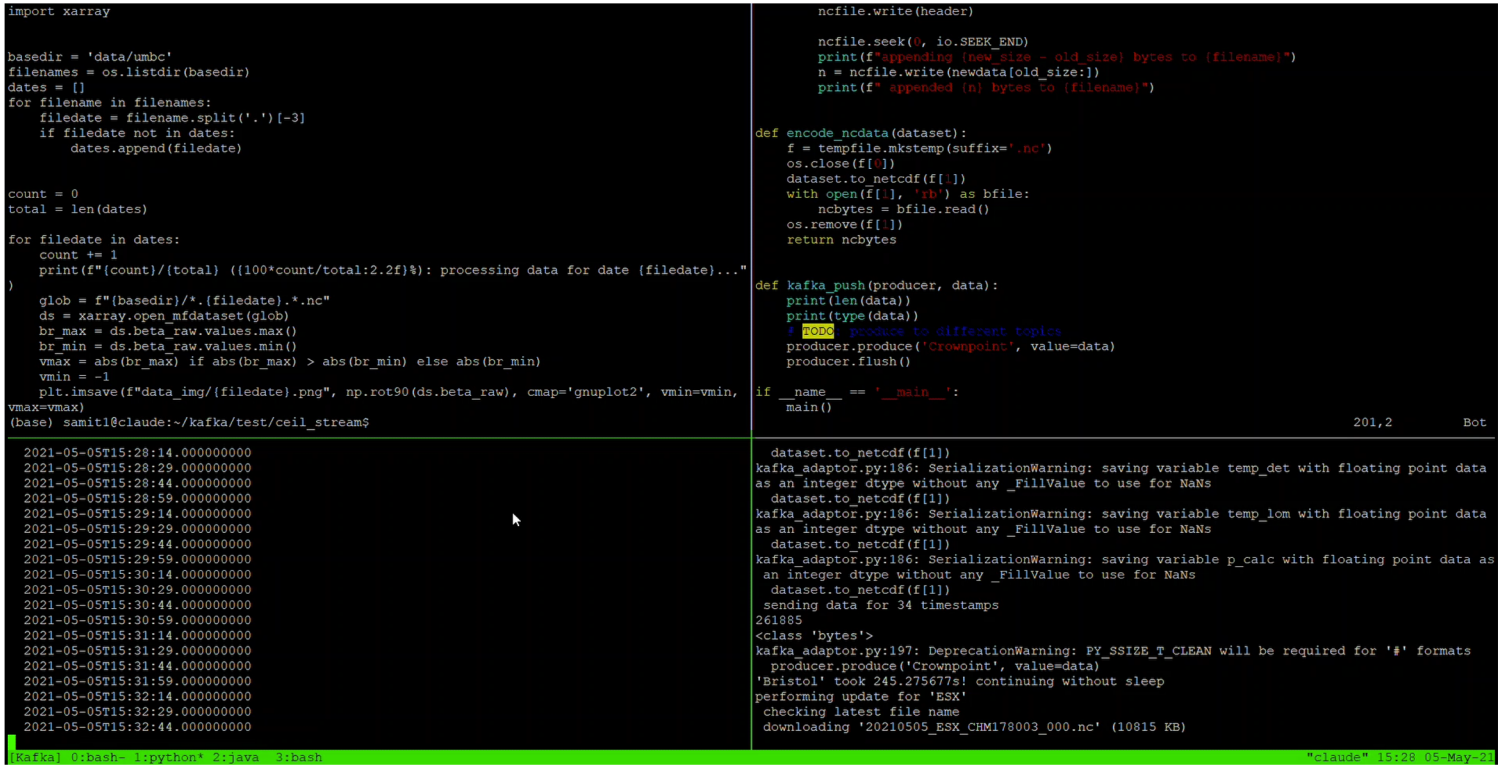




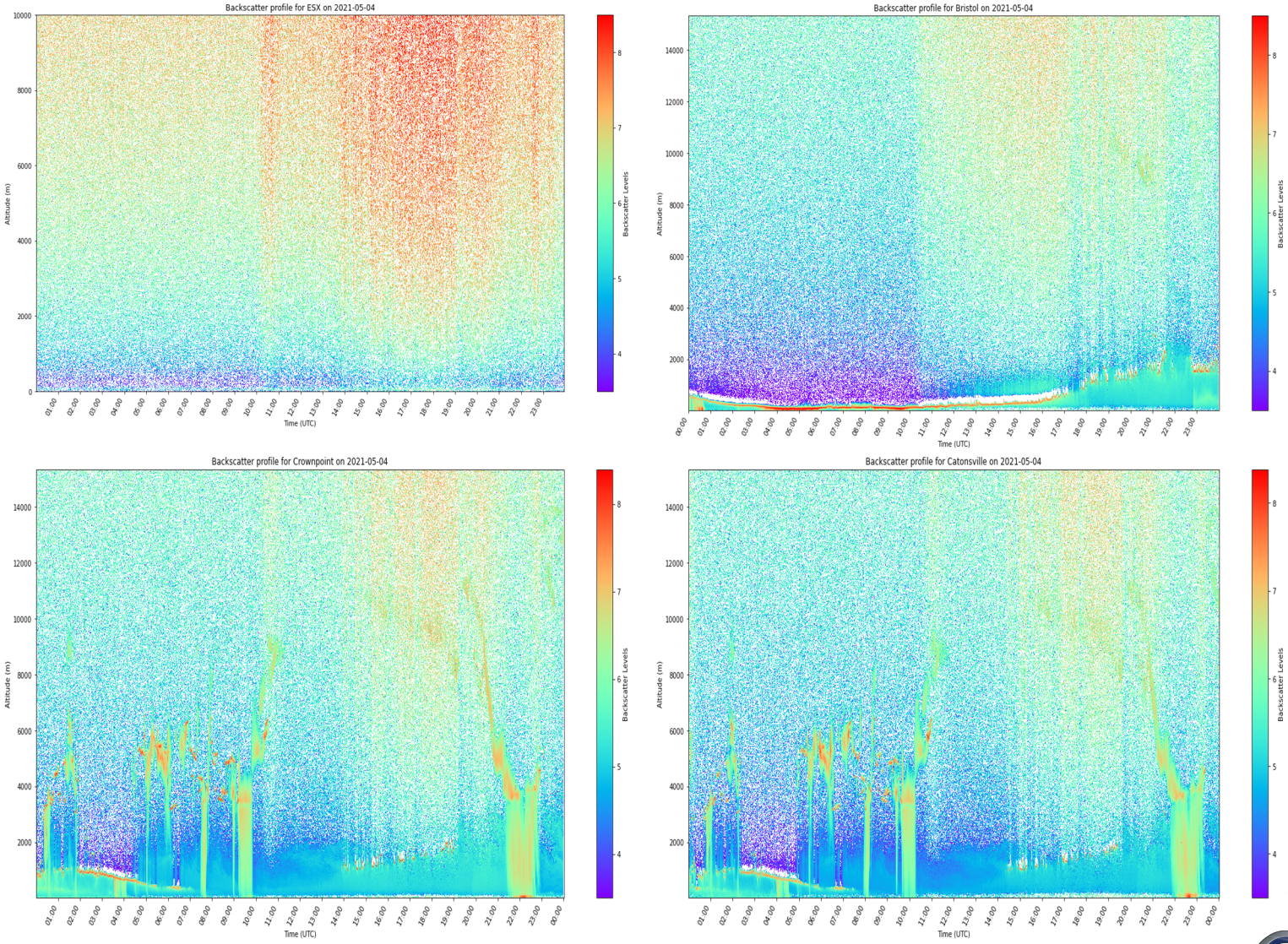
Streaming from NTU ceilometer measurement

Kafka Edge Streaming Ceilometers

- https://drive.google.com/file/d/1hRBh9C31q03jQQ8x_jf2rLvZ4HmkP22P/view?usp=drive_web



Edge Streaming Prototype: Secure, Fault Tolerant Auto Data Ingestion, Processing and Analytics



ESTO
Earth Science Technology Office

Task 2: Using Machine Learning to Identify PBL Heights

Dr's Jennifer Sleeman, Vanessa Caicedo, Dorsa Ziaei, and Zeenat Ali, and Kinjal Patel

Operational method used to estimate and predict heights given what has been learned from past PBL height identification, station location, ceilometer type, and model variables.

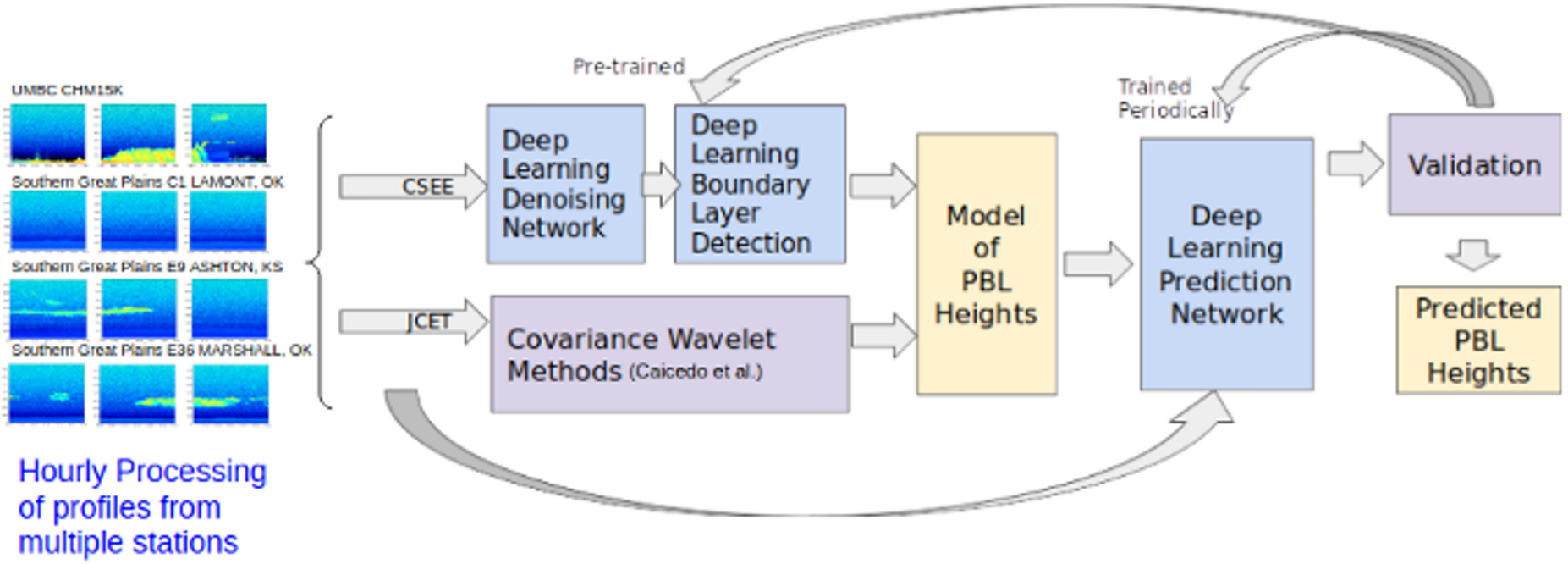
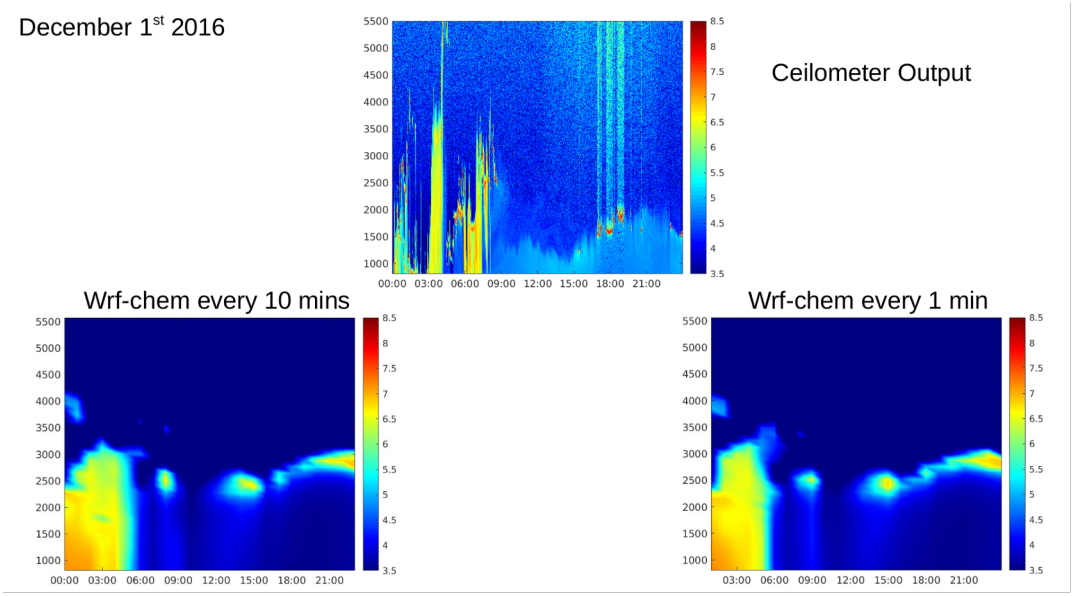


Figure 1. Integration of Machine Learning Methods for Operational PBLH

Operational System: Using Machine Learning to Identify PBL Heights
Dr's Jennifer Sleeman, Vanessa Caicedo and Dorsa Ziaei

Current Efforts and Updates:

- Continued work on Deep Boundary Layer detection as part of WRF-CHEM
- Evaluating performance for simultaneous processing of 1000's of geographical locations



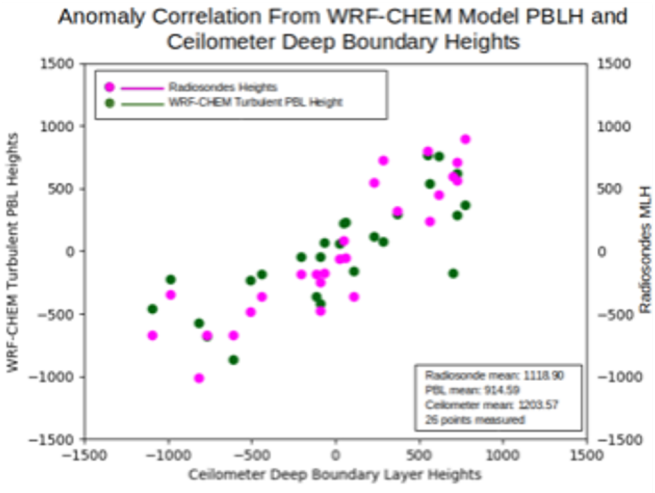


Figure 3. Comparing WRF-CHEM models with Ceilometer Output for UMBC Location. Dec 1, 2016.

Multi-Station Processing and Integration

We have processed over 11 stations now.

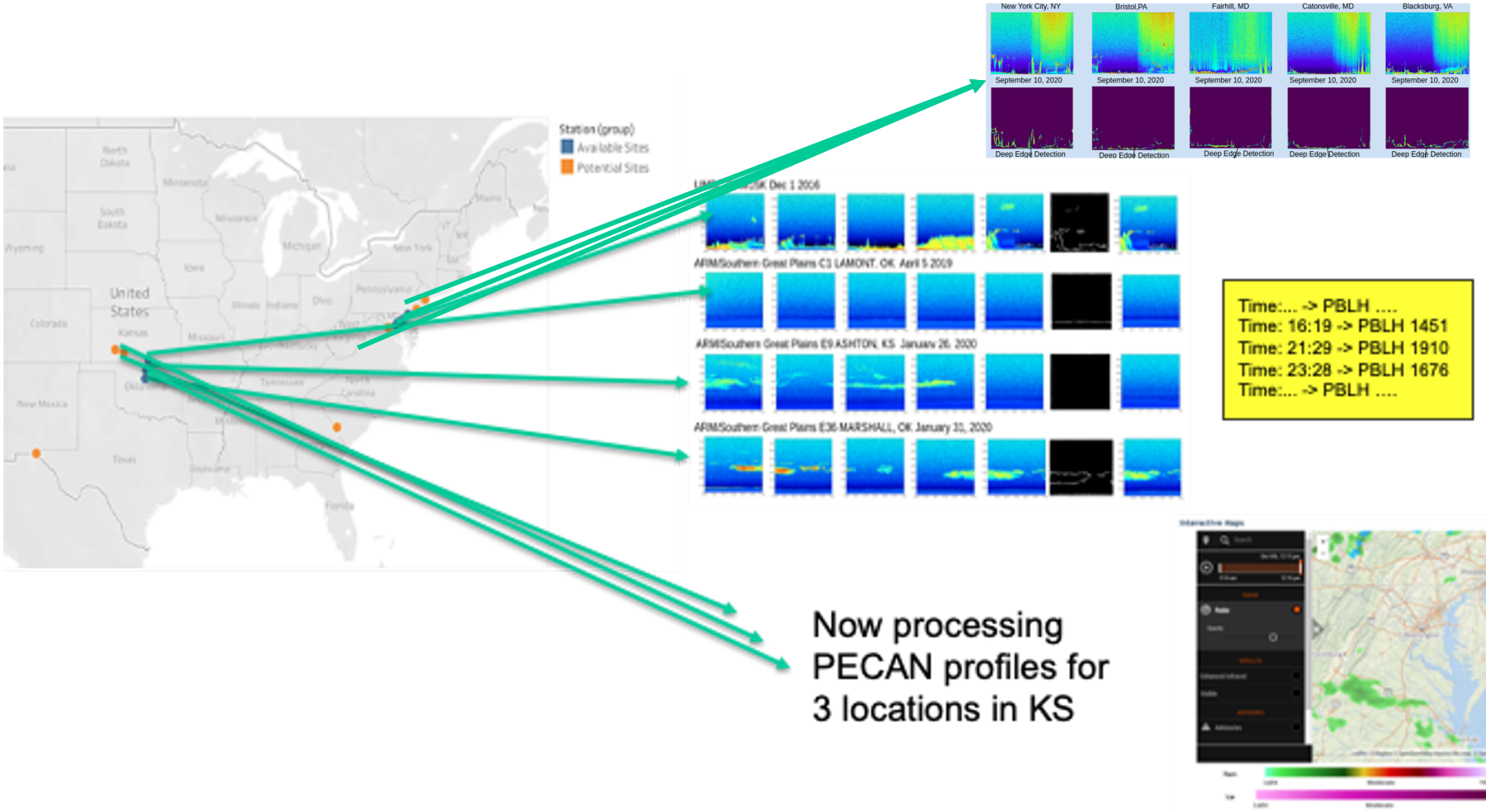


Figure 4. Operational Effort for Multi-Station Processing

Multi-Station Processing and Integration

We have integrated multi-station processing with the data acquisition team and visualization team for a September 10th experiment.

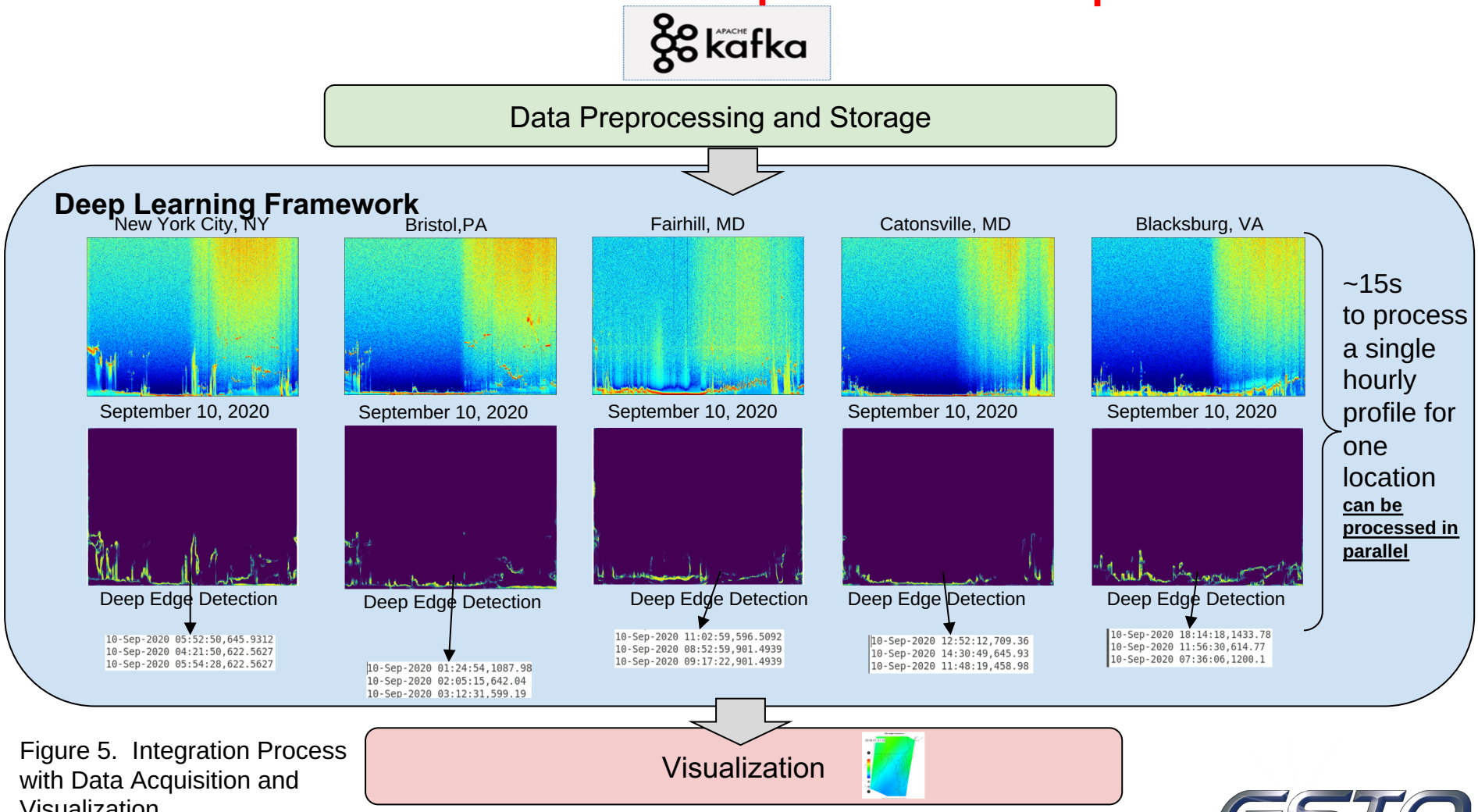


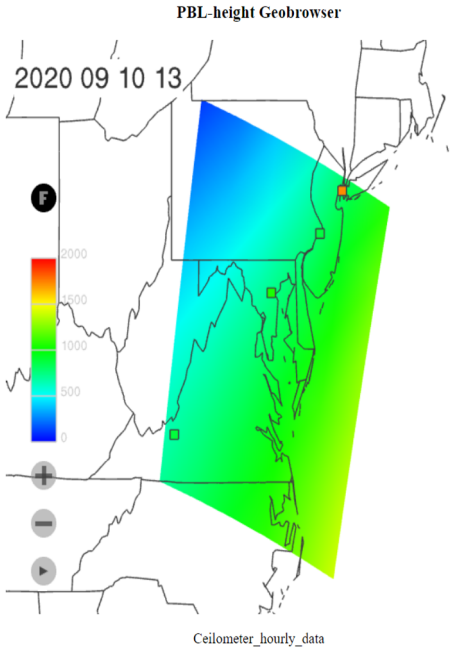
Figure 5. Integration Process with Data Acquisition and Visualization

Task 3: PBLH Hourly Spatial Fusion and Visualization

D. Chapman, P. Bindu

- Objectives:
  - Create Level 3 hourly gridded PBLH from ceilometers
  - Hourly data fusion of Ceilometer ABLH with WRF-Chem Model output PBLH
    or NOAA HRRR Model forecasts
  - Web accessible interactive visualization + geobrowser

- Compressive Sensing Fusion of PBLH from Ceilometer and WRF-Chem Model:
  - Interactive Data visualization using-
    - U.S. Census Bureau's MAF/TIGER Database
    - HTML5 web servlet technology
    - Integration with Data Archive + Apache



Data Fusion

Ceilometer PBLH

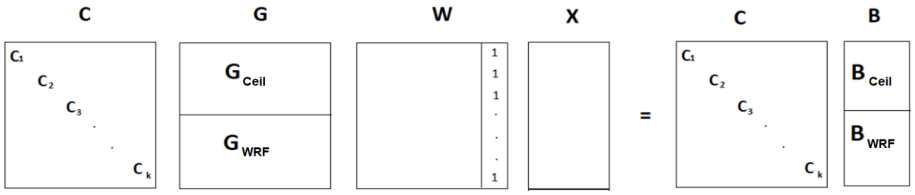
WRF-Chem PBLH

Level 3 Gridded PBLH via Compressive Sensing Fusion

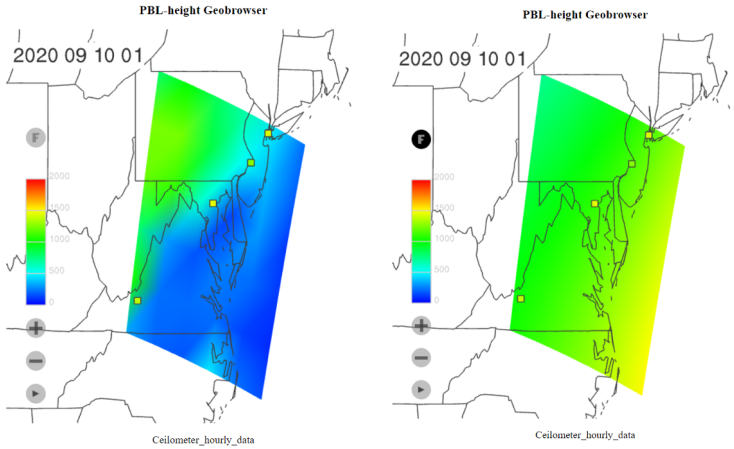
- **Compressive Sensing for Level 3 PBLH grid**
  - Fusion of Ceilometer Profiles and WRF-chem to infer gridded PBLH at 3.2km resolution.
  - Fusion with WRF-chem model outputs can interpolate between ceilometer point backscatter measurements while maintaining high frequency signal due to surface interaction.
  - PBLH spatial fusion using L1 Compressive Sensing with Wavelet basis space.

$$\min \quad ||CGWx - Cb||_2^2 + \lambda ||x||_1$$

  - PBLH profiles from 5 ceilometers along greater BW metropolitan area.

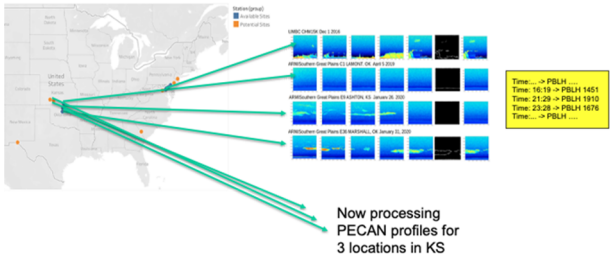


C: diagonal calibration matrix, G: Sensing matrix, W: wavelet transform
X: Inferred Signal B: Observation vector



Interactive Visualization of L3 Fused PBLH product

(Left: wrf-chem PBLH, Right: fused PBLH)



Ceilometer Stations including PCAN and greater BW metro area

ESTO Earth Science Technology Office

Summary of Accomplishments
B. Demoz, M.Halem

1. **Towards** demonstrating end-to end feasibility of hourly, edge streaming, secure, scalable, fault-tolerant **nationwide** pilot:
   Level 1B - ingest ceilometer and satellite-based Lidar aerosol backscatter profiles,
   Level 2 - pre-process and infer aerosol boundary layer heights using machine learning
   Level-3- visualize hourly fused updates using compressive sensing in near real-time.

2. **Developed** and validated Hybrid Deep Hierarchical Machine Learning Edge detection and Covariance Wavelet algorithm for end-to end hourly aerosol boundary layer height (ablh) product from Ceilometer and remote sensed Lidar aerosol backscatter and model generated backscatter simulations.

3. **Produced** 3-D hourly boundary layer height maps by a compressive sensing fusion methodology from the derived ceilometer ablh and operational reanalysis pblh.

Thanks to ESTO/AIST

Publications

2021 Rahul Gite, Samit Shivadekar, Milton Halem , Phuong Nguyen. Near Real-time Streaming and Data Processing of Ceilometer data using Kafka. Kafka-Summit 2021

2021 Phuong Nguyen, Rahul Gite, Zhifeng Yang, Halem "Deep Neural Network Architecture Search for Emulating Physical Parameterization of Planetary Boundary Layer Height" IEEE IGARSS 2021

2021 Halem M. Zhifeng Yang, C. Cruz,, J. Sleeman," Coupling AI to Aerosol Model Parameterizations for inferring Boundary Layer Heights"101st AMS 11R2O 2021

2021 Sleeman J., D. Ziaei, Z. Yang, V. Caicedo, C. Calderella, M. Halem, R. Delgado, B. Demoz, " A Deep Multi-Stacked NeuralNetwork Approach for Improved Planetary Boundary Layer Height Estimation" AMS 101st Annual Meeting 8.6

2020 Halem M., J. Sleeman, Z. Yang, M. Chin, D. Watson-Parris, B. Demoz, " Feasibility Studies of Cloud Resolving NU -WRF Subseasonal Forecasts with AI Emulations". AGU IN024 Fall 2020

2020 Gite R., M. Halem, P. Nguyen, " Compressive Sensing and Deep Learning framework for Multiple Satellite Sensor Data Fusion". AGU IN033 Fall 2020

2020 P Nguyen, R Gite, A Rathod, Z Yang, M Halem Machine Learning for emulating Physical Parameterization of Planetary Boundary Layer Height, AGU 2020

2020 Sleeman J., Vanessa Caicedo, Dorsa Ziaei, M. Halem, Belay Demoz[1]uben Delgado," Using Machine Learning to IdentifyPlanetary Boundary Layer Heights for Ceilometer-Based LIDAR Backscatter Retrievals". AGU Fall 2020

2020 Yang, Z., M.Halem " Model Evaluation and Assimilation of the Planetary Boundary Layer Height". AGU Fall 2020

2020 Nguyen P., M. Halem,'Satellite Data Fusion of Multiple Observed XCO2 using Compressive Sensing and Deep Learning' IGARSS 9/20

2020 Ayanzadeh R., M. Halem, T. Finin,"An Ensemble Approach for Compressive Sensing with Quantum Annealers" IGARSS 9/20

2020 Ziaei, D., D. Chapman, Ya. Yesha, M. Halem, "Segmentation of Stem Cell Colonies in Fluorescence Microscopy images with transfer learning" SPIE Conference Medical Imaging, 2020 Houston, TX

2020 Ayanzadeh, R, M. Halem, and T. Finin. "Reinforcement quantum annealing: A hybrid quantum learning automata." Nature: Scientific reports 10, no. 1 (2020): 1 11.

Publications cont.

2020 Carroll , B., Belay B. Demoz , David D. Turner , and Ruben Delgado: Lidar observations of a mesoscale moisture transport event impacting convection and comparison to Rapid Refresh model analysis" Published-online: 04 Dec 2020 Collections: Plain Elevated Convection At Night (PECAN)DOI: https://doi.org/10.1175/MWR-D-20-0151.1

2020 Tangborn, A., Demoz, B., Carroll, B. J., Santanello, J., and Anderson, J. L. "Assimilation of lidar planetary boundary layer height observations, Atmos. Meas. Tech. https://doi.org/10.5194/amt-2020-238

2020 Lopez-Coto, Israel; Micheal Hicks; Anna Karion; Ricardo Sakai; Belay Demoz; Kuldeep Prasad; James Whetstone: assessment of Planetary Boundary Layer parameterizations and urban heat island comparison: Impacts and implications for tracer transport J. Appl. Meteor. Climatol. (2020) 59 (10): 1637–1653.https://doi.org/10.1175/JAMC-D-19-0168.1

2019 Nguyen, P, M. Halem,"Deep Learning Models for Predicting C Employing Multivariate Time Series" IEEE Knowledge and Data Discover (KDD) conference Aug. 2019, Anchorage, Alaksa.

Real-time Demonstration Using Smoke data from Sept. 17 2020

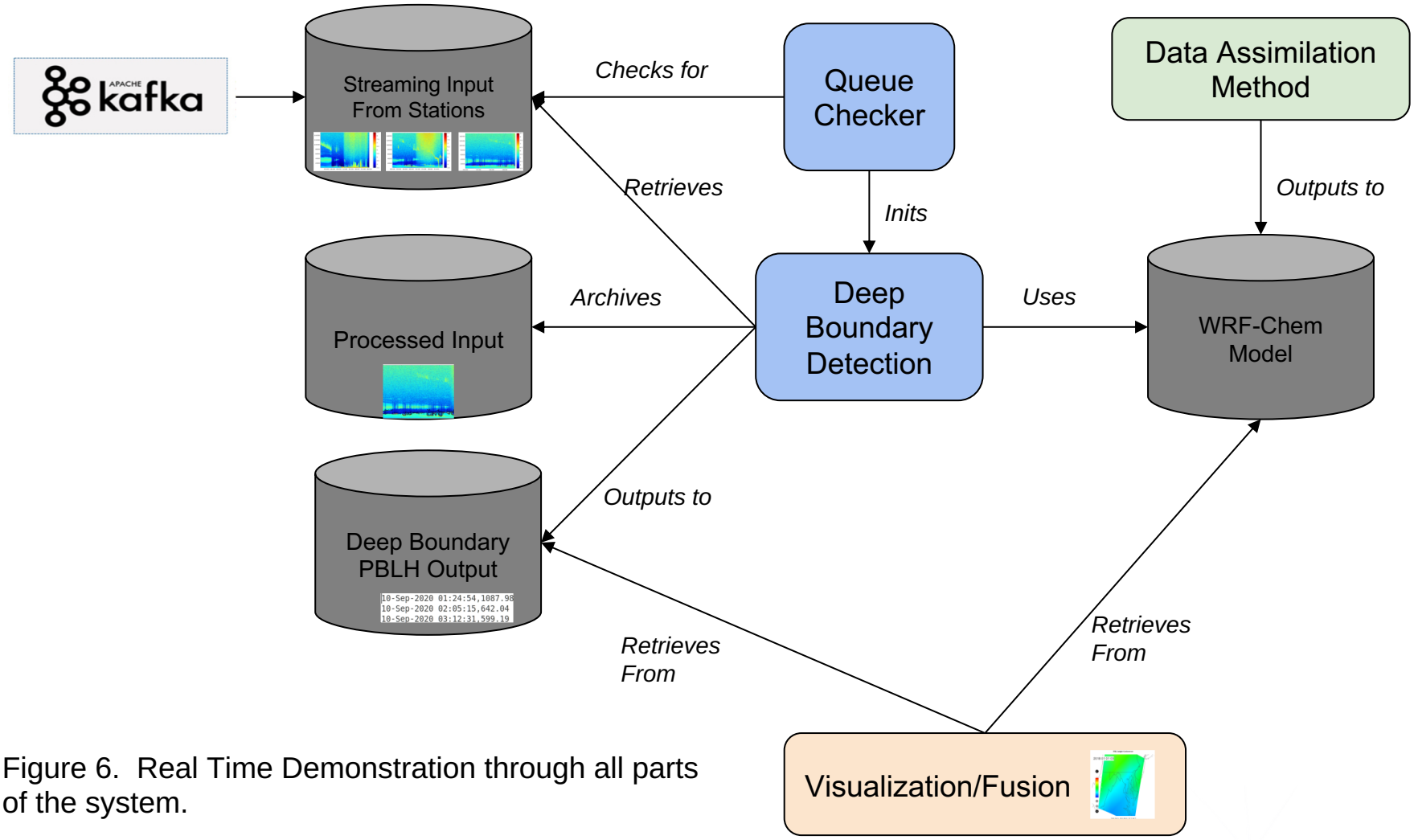


Figure 6. Real Time Demonstration through all parts of the system.

California Wildfire Smoke
September 17, 2020

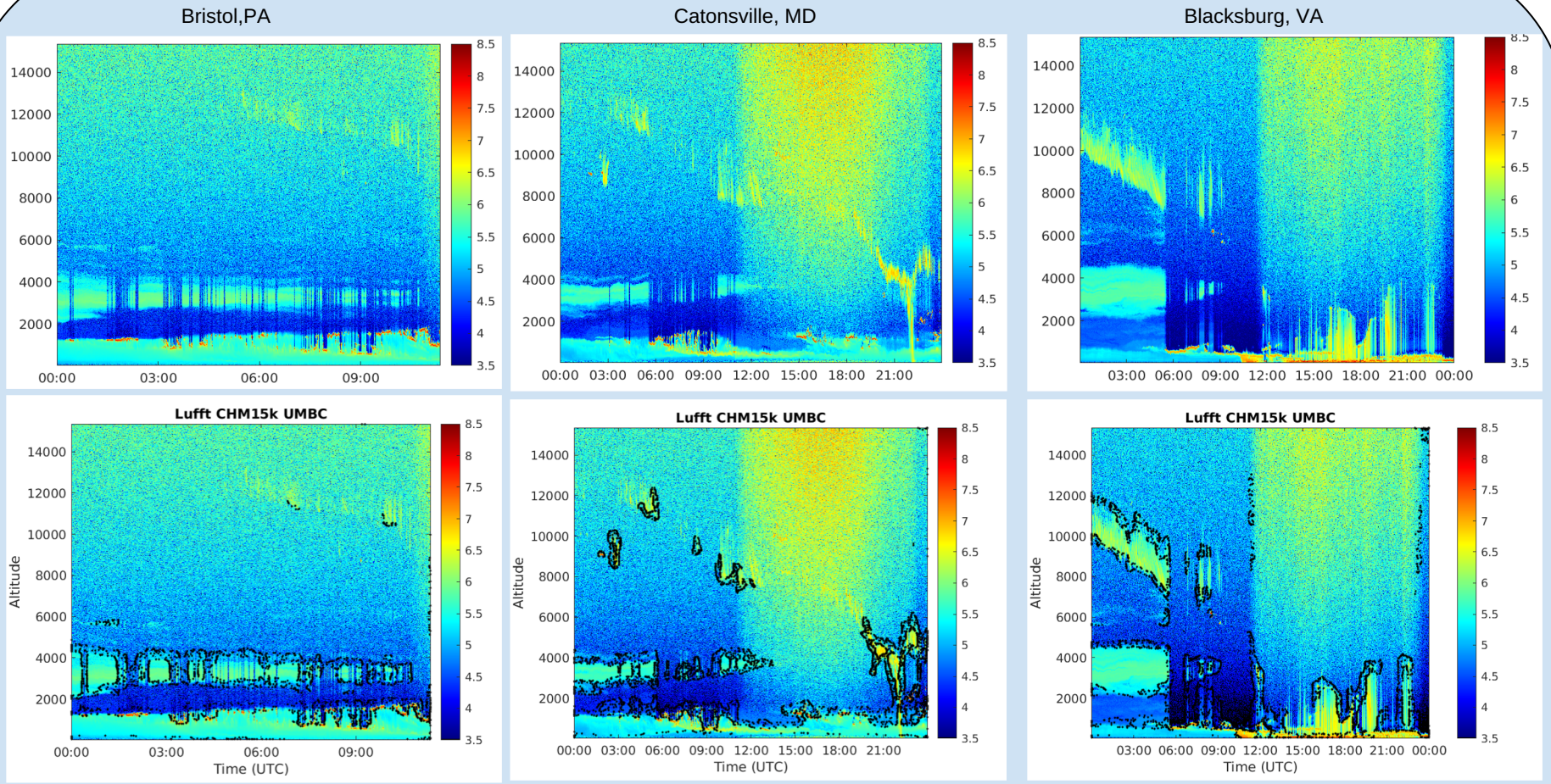


Figure 7. Ceilometer Output for September 17, 2020 and Deep Boundary Layer Applied (Points in Black)

ESTO Earth Science Technology Office

Extending Deep Boundary Detection to Satellites : ICESat-2

- Current efforts underway to compare a traditional method for estimating PBLH for ICESat2 data and using the Deep Boundary Layer Detection method
- Stacked LSTM to process ICESat2 data (working in combination with the WRF-CHEM model data LSTM and Ceilometer-based LSTM)
- Results will be forthcoming in a future meeting

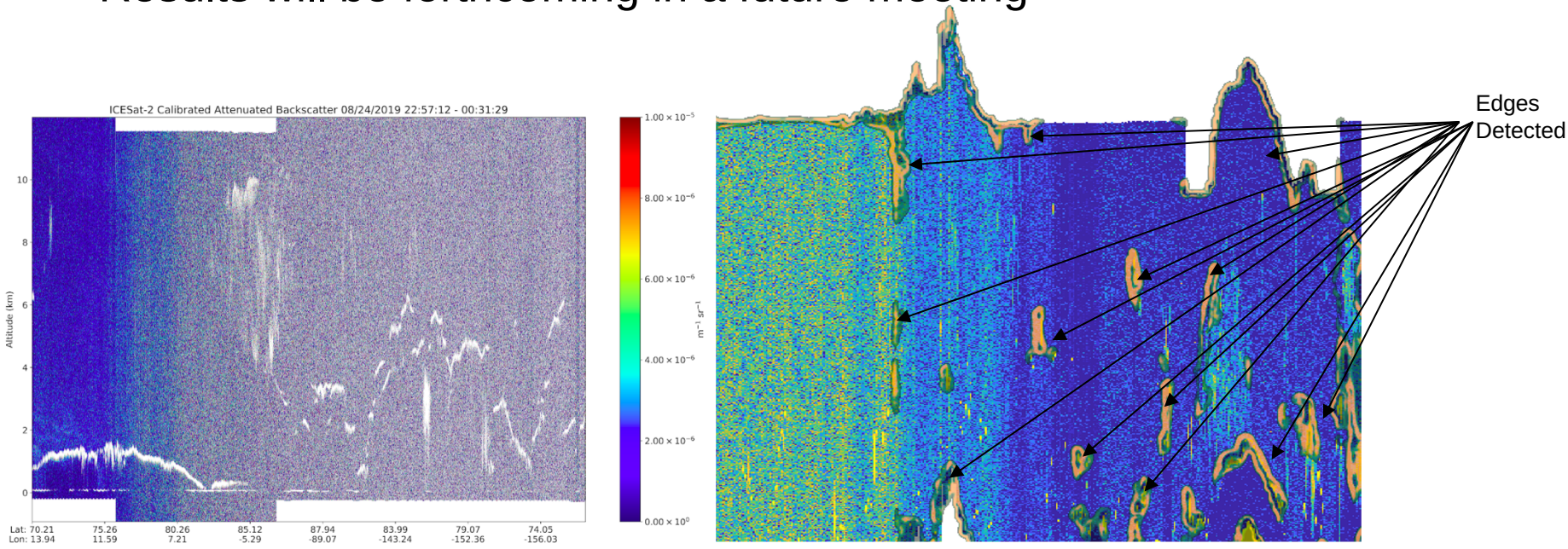


Figure 14. ICESat2 for Arctic and 6b. ICESat2 -ATL09_20190501105015_05070301_003 With Overlay of Edges Detected for Location around UMBC

A Deep Learning Ceilometer (LIDAR)-based Atmospheric Boundary Layer Height Product Over CONUS

PI: M. Halem, CO-PI:B. Demoz, UMBC

Objectives:

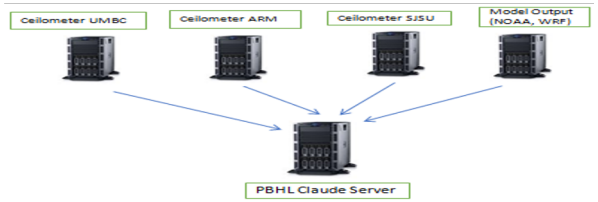
Task 1: Identify, acquire and implement an internet, **hourly edge streaming, secure, fault-tolerant ingest L1 system of Ceilometer,** Satellite and Model- based LIDAR backscatter observations over the CONUS to generate L2 ABLH products.

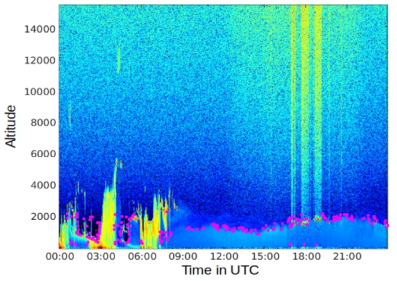
Task 2: Develop and **test automated synchronized hybrid L2 ABLH** LIDAR processing system for continental wide US profiles combining Machine Learning, Wavelets and Mixture of Experts to generate hourly product with validating error bounds.

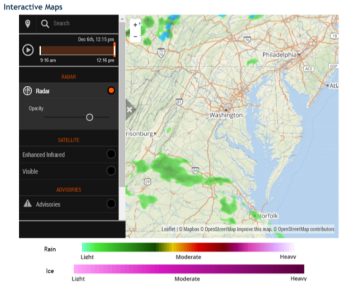
Task 3: Generate regional and CONUS wide L3 hourly visualizations and longer-term animation products. Provide data management, archival and community delivery system of LIDAR Level 1, 2 and 3 products

Task 4: Conduct model data assimilation of ABLH output and radiosonde acquisition system for product validation and verifications.

Task 5. Produce quarterly reports and conduct semi-annual reviews and convene external advisory group for system evaluation.







Approach

- Data Acquisition plan.
  Integrate 4 JCET +3 CSEE ceilometers into automatic data ingest system
- Develop a hybrid machine learning processing system for generating hourly ABLH. Provide Project ATBD or on Github for processing system.
- Validate v1.0 performance and accuracies during op'ns test.
  Identify areas for Improving edge detection method. Continue evaluation of v1.0 methods Add denoising method
  Integrate the LSTM method with the boundary detection method.
- ABLH spatial Visualization.
  Create ABLH spatial maps and dynamic visualizations.
  Fuse UMBC hrly ceilometer ABLH with NOAA PBLH forecast.
  Obtain external evaluation feedback and update visualization

4/09

Key Milestones

-Acquire 3 NASA Luft ceilometers, install at VA Tech, Bristol PA and NTU ceilometers and conduct 1st end-to-end system test of (10) edge streaming ground system Level 1/2/3 operations. 6/21

-Conduct 2nd level 1/2/3 end-end test with ground/satellite and model generated backscatter data in near real time. 9/21

-Produce a robust Ceilometer web-based ABLH hybrid machine learning based system scalble to processing streaming 5-minute data from more than 100 ceilometer stations. 11/21

- Provide a visualization service of PBLH products and generate spatial hourly plots with Zoom capabilities 9/21

- Demonstrate fault tolerant, secure, edge streaming 1- week end-to end validated test of the unified hybrid ground/space/model AI/ML generation of Regional ABLH web accessible surface 11/21

TRL_{in}= 3 **TRL_{fin} = 6**

ESTO Earth Science Technology Office

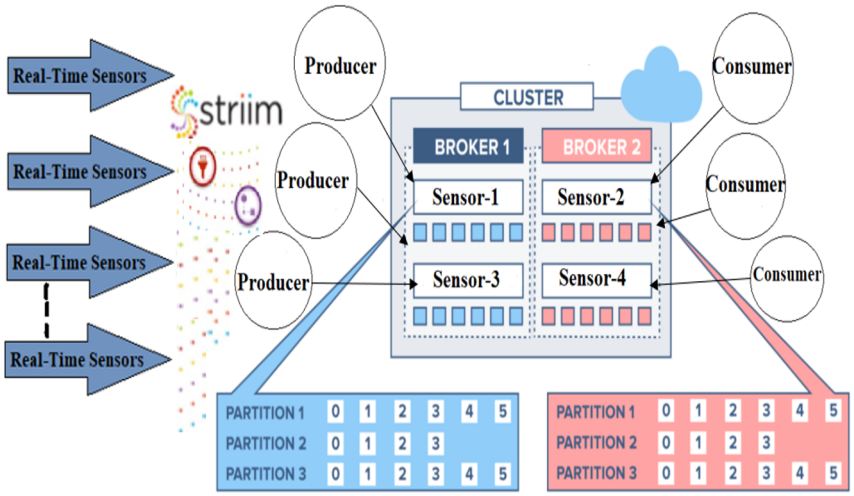
Edge Streaming Prototype: Secure, Fault Tolerant Auto Data Ingestion, Processing and Analytics

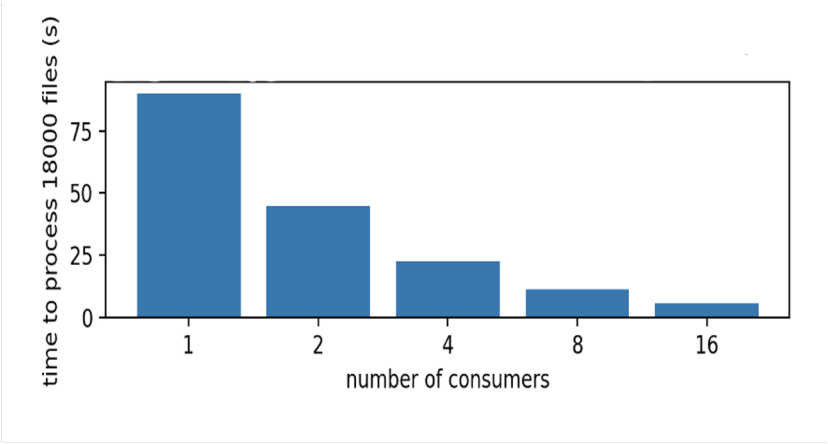
- **Kafka run using distributed cluster of GPU Servers: to** train AI models and pipeline processing

- duplicate historical UMBC Ceilometer's 5 minute observations to measure latency and improves throughput and scalability performance. compute test (LSTM inferences using backscatter radiation)

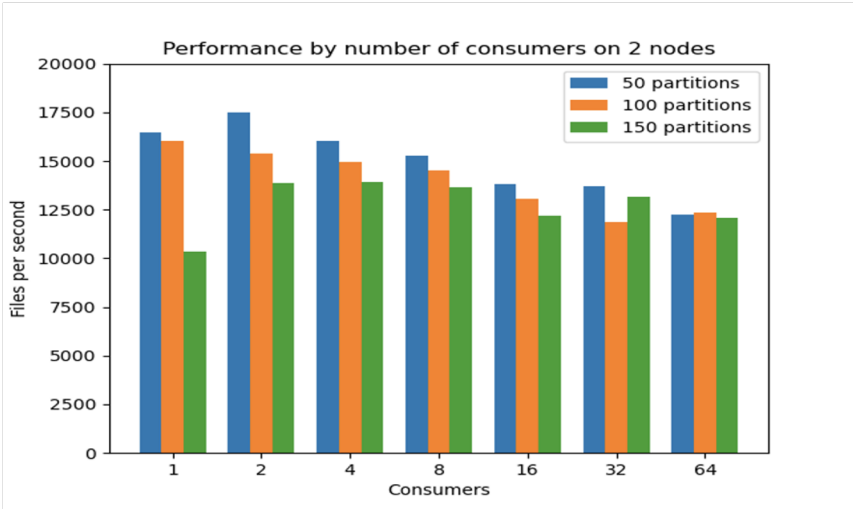
⇒ scale to process thousands of files per second.

⇒ all consumers are processed files in parallel. each consumer handle 50-150 partitions in parallel





Latency Single node (simulated speed for 18000 files (5 minute data) at 200.0 files per second).



Throughput performance when two broker are present across a cluster of two nodes.

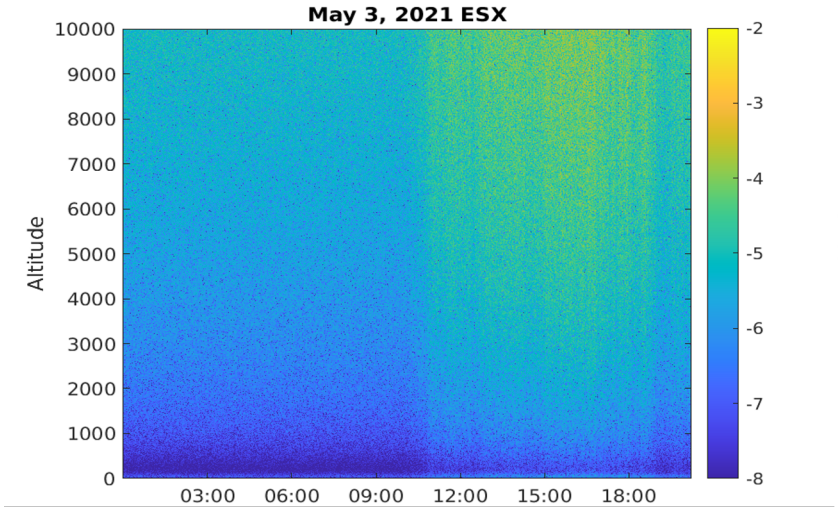
ESTO Earth Science Technology Office

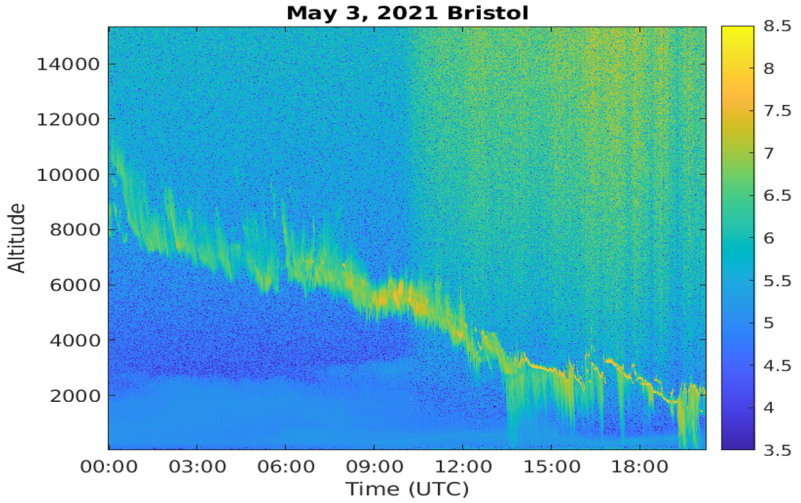
PBLH Level 3 Spatial Fusion and Visualization

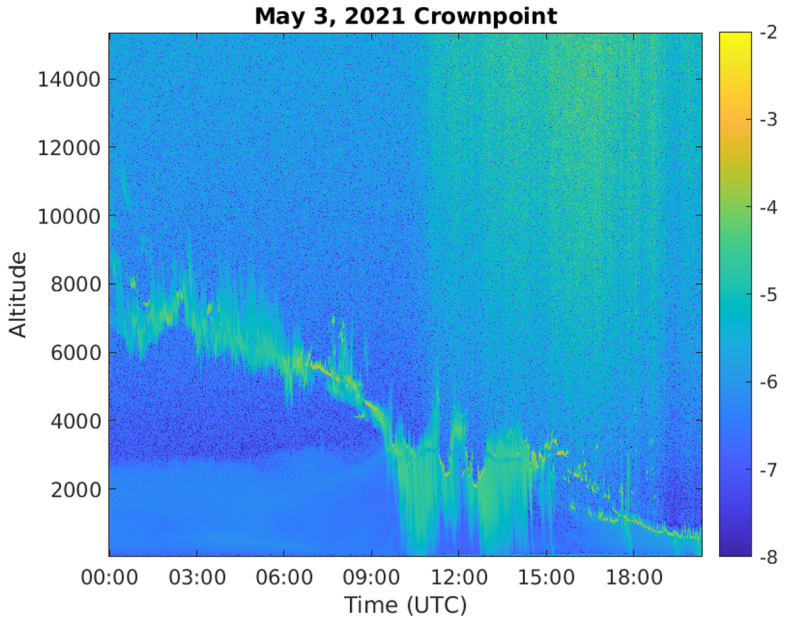
Key Accomplishments

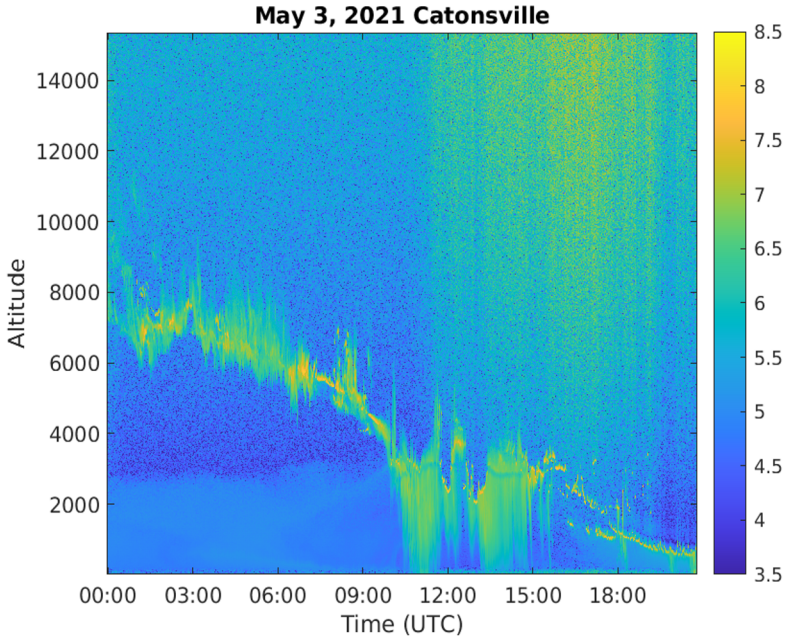
- Integration of Ceilometer profiles from four sites along i95 corridor
  - End to end processing including data streaming, L2 retrievals and L3 gridded PBLH maps
- Fusion of L2 ceilometer profiles with WRF-CHEM model outputs
  - Method of compressive sensing with 2D+time wavelet transform
  - 3.2km resolution and hourly timescales over BW i95 corridor
- Interactive visualization geobrowser
  - Display ceilometer derived L3 gridded PBLH profiles
  - Comparison of L3 product with raw WRF-chem shows large differences in PBLH

Edge Streaming Prototype: Secure, Fault Tolerant Auto Data Ingestion, Processing and Analytics





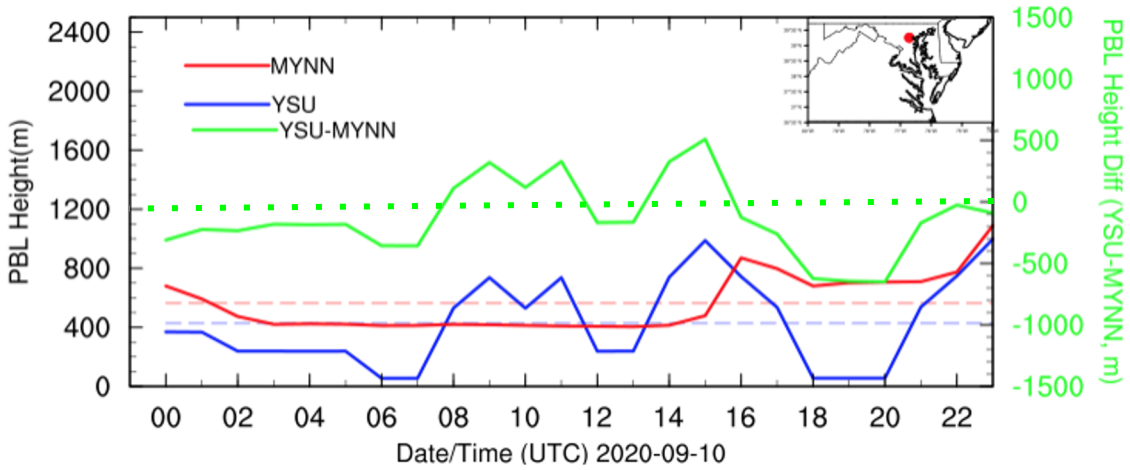




Earth Science Technology Office

Variability of Diurnal Variation PBLH to parameterization

Sep 10, 2020



@Catonsville:

- Diurnal cycle;
- YSU: fluctuation; MYNN: smooth
- Afternoon&Nighttime: $PBLH_{YSU} < PBLH_{MYNN}$; Morning: $PBLH_{YSU} > PBLH_{MYNN}$;
- $PBLH_{YSU} = 0$?

ESTO Earth Science Technology Office

Presentation Overview

- Preliminary results of our end-end test of an hourly Atmospheric Boundary Layer Height (ABLH) product from a deep machine learning Lidar based ceilometer and satellite aerosol backscatter data.
- Deployment of an Edge Streaming Aerosol Backscatter Cyberinfrastructure Ecosystem.
- Emulation of WRF Microphysics on Thermal Structure using Auto Keras Neural Architecture Search.
- Hourly ABLH Simultaneous Processing of Multi-Ceilometer Sites Using a Hierarchical Machine Learning Edge Detection and Denoising Neural Network.
- Preliminary Results Extending ABLH Hourly Processing System to NU-WRF-CHEM backscatter with GFS forcing and real-time fire smoke emissions.
- Towards Processing Multi-source Aerosol Backscatter profiles from Ground, Space and Model using Convolutional LSTM.
- Spatial- Temporal Fused Regional Hourly ABLH and WRF/GFS Visualization and Animation.

Next Steps:

(i) Test larger distributed ceilometer system (12); merging with satellites (Icesat-2/ADM Aeolus) Lidars

(ii) Include block chain, enhanced fault tolerance, Tensor Edge Streaming Decomposition

(iii) Evaluate NU-WRF-CHEM ABLH Data Assimilation, Train a NAS[1] AI Microphysics Emulator

(iv) Develop a Hi-res Progressive Neural Net; Embed Deep HED[2] in GOCART/Microphysics

[1] Neural Architecture Search

[2] Hierarchical Edge Detector